Schank, R. C. (1990). *Tell me a story: A new look at real and artificial memory.* New York: Scribner's.
Schon, D. (1990). *The theory of inquiry: Dewey's legacy to education.* Presented at the second annual meeting of the American Educational Research Association, Boston, MA.
Singer, J. L. (1980). The power and limitations of television: A cognitive affective analysis. In P. H. Tannenbaum & R. Abeles (Eds.), *The entertainment functions of television* (pp. 31–65). Hillsdale, NJ: Lawrence Erlbaum Associates.
Spohrer, J. (1999). Information in places. *IBM Systems Journal, 38*(4).
Thorndyke, P. W., & Yekovich, F. R. (1979). *A critique of schemata as a theory of human story memory* (P-6307). Santa Monica, CA: Rand.
Zimmerman, B. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202–226). New York: Cambridge University Press.

Chapter 4

Evidence for Learning From Technology-Assisted Instruction

J. D. Fletcher
Institute for Defense Analyses

Should we use computer-based technology to teach? What evidence is there that doing so produces, assists, or promotes learning? What evidence is there that learning occurs better (i.e., more effectively or efficiently) with technology? Are the benefits of technology-assisted instruction worth whatever must be given up to get them?

These questions do not yield facile answers, but they seem timely and appropriate given the current state of the art. We have been applying technology in learning long enough to expect to find evidence suggesting that these efforts either are or are not worth pursuing. In fact much such evidence has been presented in the research literature. This chapter attempts to summarize this evidence and the case it makes for the use of technology (specifically, computer-based technology) in instruction.

Contrary to some casual understanding, research, development, use, and assessment of computer technology in the teaching–learning process did not begin with the introduction of personal computing in the late-1970s. Such work began much earlier.

For instance, the University of Illinois Coordinated Research Laboratory began developing what Chalmers Sherwin in the mid-1950s called "a workbook with feedback" (Bitzer, Braunfeld, & Lichtenberger, 1962). The development became the well-known Programed Logic for Automated Teaching Operations (PLATO) project, which was intended to unleash the creative instructional energies of professors and other instructional personnel who would proceed to produce and implement PLATO lessons for their classes. At about the same time, the IBM Research Center was supporting research and development of programs to teach binary arithmetic, stenotypy, psychological statistics, and German reading (Uttal, 1962). Efforts to integrate computer-assisted instruction with higher education (Holland, 1959) and elementary school education (Porter, 1959) were begun at Harvard University. Similar efforts were initiated at Stanford University to apply and help verify research findings from mathematical psychology, cognitive psychology, and psycholinguistics by incorporating them in elementary school mathematics, beginning reading, college-level Russian, and mathematical logic programs of instruction (Atkinson & Wilson, 1969; Fletcher, 1979; Suppes, 1964). Other early efforts, such as those at the University of Texas, U.S. Naval Academy, University of Southern California, Bolt, Beranek, and Newman, Inc., Air Force Personnel and Training Research Center, and Pennsylvania State University, could also be listed.

In short, there have been more than 45 years of research, development, use, and assessment of computer applications in instruction. By now, we should have some idea of the promise offered by such applications. We should know whether investment is warranted for further research and development or even if these applications are ready for large-scale implementation.

WHAT INSTRUCTIONAL TECHNOLOGY ARE WE TALKING ABOUT?

It may be a good idea to identify, roughly, what we are talking about. Basically, the topic concerns interactive instruction that is tailored on demand to the needs of individual learners. Much of this instruction depends on computer technology for delivery and presentation and includes such applications as computer-based instruction, interactive multimedia instruction, "intelligent" tutoring systems, networked tutorial simulation, and web-based instruction.

Today's emphasis on distributed learning and instruction leads us to describe these technologies as asynchronous because they can deliver instruction and mentoring (problem solving, performance aiding, and support) anywhere and at anytime. These technologies are frequently used in residential classroom settings, but they do not require learners to gather at specific times and specific places in order to learn. There may be (and often is) a (human) instructor in the equation, but there is no requirement that the instructor and the students be actively engaged at the same time.

One technology for asynchronous instruction should not be forgotten. This technology is remarkable for its easy portability and ubiquity, its random access to text, graphics, and color, its modest environmental requirements, its minimal power requirements, and, especially, its low cost. It has been available and in use for over 500 years. It is, of course, the technology of books.

Gutenberg's development of movable, metallic type may be seen as the second of three major revolutions in instruction. First was the development of writing. Before writing, the content of high quality instruction was available only synchronously, face to face, as instructors engaged their students in learning. Writing made the content of instruction available anytime, anywhere. Gutenberg's press and the books it printed effected a second revolution in instruction by making the content of high quality instruction widely, asynchronously, and (eventually) inexpensively accessible.

A third revolution in instruction may now be occurring. It is founded on the rapid and continuing development of computer technology. The computer-based instructional technologies listed above make both the content and the interactions, the tutorial give and take, of high quality instruction widely, asynchronously, and inexpensively accessible. They are available anytime and anywhere, and they can initiate relevant and appropriate instructional interactions on their own. They can be designed to adapt and respond to the needs and intentions of individual learners. They may foment a third revolution in instruction that is at least as significant as the first two.

These new and emerging instructional technologies will affect what, how, where, when, and for whom instruction and learning take place. They can provide expert mentors, tutors, guides, and coaches tailored to every learner or groups of learners. They can provide decision aiding and problem-solving assistance as well as instruction. They can adjust the pace, content, difficulty, and sequencing of instructional material and its presentation to individuals or groups of individuals in accord with their needs and intentions. If we use them well, they can become the foundation for a society of lifelong learners who can contend with rapid changes in work, commerce, and daily living, in other words, learners who function and compete successfully in a rapidly changing, technology-driven world economy.

INDIVIDUALIZATION AND INTERACTIVITY IN INSTRUCTION

Many intuitively appealing arguments have been made for the use of computer technology in instruction. They include immediate reinforcement of student responses, privacy of responses, culture-free fairness, infinite patience, and sensory immersion. All these capabilities have been verified to some extent by empirical research. But where is the leverage? What matters most and makes the most difference? Individualization seems to be the key.

The principal payoff from application of computer technology in instruction may come from their capabilities to tailor highly interactive environments to the needs of each individual learner.

The value of these capabilities may be seen in comparisons of one-on-one tutoring (one instructor with one student) with one-on-many classroom instruction (one instructor with 20–30 students). Benjamin Bloom (1984) and his students at the University of Chicago completed three such comparisons. Such a difference in instructional presentation might be expected to favor one-on-one teaching. What is surprising is how much it matters. Across these studies, the difference in student achievement amounted to two standard deviations. This difference is roughly equivalent to raising the achievement of 50th percentile students to the 98th level of achievement.

Some reasons for this difference are suggested by other research comparing one-on-one tutoring with one-on-many classroom instruction. Consider, for instance, some findings on the time it takes for different students to reach the same instructional objectives:

- Ratio of time needed by individual kindergarten students to build words from letters: 13 to 1 (Suppes, 1964)
- Ratio of time needed by individual hearing-impaired and Native American students to reach mathematics objectives: 4 to 1 (Suppes, Fletcher, & Zanotti, 1975, 1976)
- Overall ratio of time needed by individual students to learn in grades K–8: 5 to 1 (Gettinger, 1984)

This classroom diversity presents a daunting challenge to teachers using current methods of instruction. How can they ensure that every student has enough time to reach given instructional objectives? At the same time, how can they allow students who are ready to do so surge ahead? How can they cope with this variability? The answer, of course and despite heroic efforts to the contrary, is that they cannot. Most classrooms contain many students who, at one end of the spectrum, are bored and, at the other end, are overwhelmed and lost. By contrast, one-on-one tutoring allows teachers to adjust pace and content to the needs and abilities of individual students. They can proceed as rapidly or as slowly as needed. They can skip what individual students have mastered and concentrate on what they haven't.

This difference in time to reach given objectives seems initially due to ability, but prior knowledge very quickly takes over (Tobias, 1982). Despite efforts to sustain common levels of knowledge in classrooms, current school practices appear to increase these differences by about 1 year for every year students spend in elementary school (Heuston, 1997). For instance, the average spread of academic achievement in grade three is about 3 years. By grade six, it increases to about 6 years.

Another factor that may explain the difference in outcomes between one-on-one tutorials and one-on-many classroom settings is the intensity of the instruction. The interactivity of one-on-one instruction compared to classroom interactivity has been studied by Graesser and Person (1994). They compared instruction using one-on-one tutoring with classroom practice in two curriculum areas: research methods for college undergraduates and algebra for seventh graders. They found the following:

- Average number of questions teachers ask a class in a classroom hour: 3.0
- Average number of questions asked by any one student during a classroom hour: 0.1
- Average number of questions asked by a student and answered by a tutor during a tutorial hour:
  Research methods: 21.1
  Algebra: 32.2
- Average number of questions asked by a tutor and answered by a student during a tutorial hour:
  Research methods: 117.2
  Algebra: 146.4

How could a classroom teacher answer 20–30 questions asked by each of 25–30 students each hour? How could a classroom teacher prepare and present during each hour of instruction, 115–145 questions adapted to the needs of each student and then provide feedback for every answer received? These considerations are obviously rhetorical. Classroom instructional practice cannot compete with tutorial instruction in the interactivity and adaptability of the instruction provided.

Why not provide one-on-one tutorial instruction to all students? The answer is obvious. We can't afford it. One-on-one instruction may be an instructional imperative, as Scriven (1975) suggested, but it is also an economic impossibility. The result is, as Bloom stated, a 2-Sigma problem. How can we, in real instructional practice, fill the two standard deviation (the 2-Sigma) gap between classroom and tutorial instruction?

Enter computer technology. Because computers can interact with learners and tailor instructional presentations to their needs, and because the computer capabilities needed to do this are less expensive than human tutors or expert consultants, some of the gap between one-on-one and one-on-many instruction can be filled, affordably, by computer technology. Computers allow us to substitute the capital of technology for the labor of human instructors. They may effect a third revolution in instruction by making individualized instructional interactions affordable.

SOME FINDINGS

Does this substitution of computers for human tutoring work? Can it fill Bloom's 2-Sigma gap? Can it create environments in which learning occurs more effectively and/or more efficiently? The following section summarizes results that have emerged from research on the effectiveness of technology applied to the problems and processes of instruction.

Technology Can Be Used to Teach

A number of studies have compared technology-based instruction to simply doing nothing. These studies did not seek to determine if these applications are a good way to teach or if they teach the right things, but simply to see if they teach anything at all.

Their findings suggest that they do. For instance, single studies by Crotty (1984) and Verano (1987) and two studies by Allan (1989) compared applications of interactive multimedia instruction with placebo treatments in which no instructional material was presented. The average effect size for these studies was 1.38 standard deviations, suggesting, roughly, an average improvement in student achievement from the 50th to 92nd percentile performance.

Significant evidence that avoids "horse race" (control group vs. experimental group) comparisons comes from early studies by Suppes, Fletcher, and Zanotti, who used computer-based instruction to provide mathematics instruction to 69 Native American (1975) and 297 hearing-impaired (1976) students. These studies developed a model for the progress, or "trajectory," of each student using CBI to learn mathematics as measured by scores on standard tests. The studies used no instructional input other than the amount of time (time on task) students spent in the CBI curriculum to predict their achievement.

The investigators found that from the 20th to 39th sessions (about 240 minutes) of computer-based instruction, they could predict to the nearest tenth of a grade placement, the score that each student would obtain on a standardized measure of total mathematics achievement based on total time in the instruction. Different students, of course, began at different levels and progressed at different rates. But different goals could be set for different students if instructional time was held constant, or different amounts of time could be assigned to different students to ensure that threshold levels of learning (as measured by the tests) were achieved. If time spent in the curriculum had no effect, no predictions would have been possible. In these studies, the precision of the predictions was as notable as the fact that they could be made and validated at all.

Technology Can Be Used to Increase Instructional Effectiveness

The more interesting question is, of course, whether technology-based instruction is an improvement: Does it allow us to create learning environments that are any better than those we already have? A typical study addressing this issue compares an approach using technology, such as computer-based instruction or interactive multimedia instruction, with what may be called "conventional instruction," using lecture, text-based materials (perhaps including programed text, and/or laboratory, hands-on experience with real equipment). There have been many such evaluative comparisons made. The findings from these studies provide a picture of what has been learned.

Early studies of effectiveness used a *box-score approach.* Investigators would determine the proportion of evaluations in which experimental group means exceeded control group means by some statistically significant extent and then report the experimental treatment as favorable or not depending on whether this proportion was large or small.

Vinsonhaler and Bass (1972) used a box-score approach to review over 30 published, empirical evaluations of computer-based instruction involving more than 10,000 students. For those who view technology-assisted instruction as a recent phenomenon, the early date of this review is worth noting. The authors found positive or equal results in 30 of the 34 studies they included for review. They found a median achievement gain among the CBI students of 40%.

Hedges and Olkin (1980) showed that the box-score approach has very low power (low ability to detect statistically significant differences) for the treatment effect sizes and sample sizes characteristic of instructional research. They also showed that the power of the box-score approach decreases as the number of studies included in the review increases.

Today, reviewers are likely to use an analysis of analyses, or *meta-analysis* (Glass, 1976). Meta-analysis allows reviewers to aggregate the results of many studies that attempt to answer a common question, such as the effect of using technology-based instruction, and report the magnitude of this effect in units of standard deviations using a common measure of merit called *effect size.*

The main drawback in using effect sizes is that they are, basically, a measure of standard deviations and not especially meaningful to individuals who are not statisticians. For this reason, the effect sizes reported here are accompanied by rough translations to percentiles. They show that an effect size of, say, 0.50 is roughly equivalent to raising the performance of 50th percentile students to that of 69th percentile students.

For instance, Kulik (1994) performed many such studies for technology-based instruction. From his own work and that of his colleagues, he reported an overall effect size of 0.35, which is roughly equivalent to raising the achievement of 50th percentile students to that of 64th percentile students by exposing them to technology-based instruction.

Examination of results from meta-analyses concerning the use of technology-based instruction, tracking improvements in both technology and instructional approaches, produces the results shown in Fig. 4.1, which suggest progress toward Bloom's targeted effect size of two standard deviations.

In Fig. 4.1, computer-based instruction summarizes results from 233 studies that involved straightforward use of computer presentations using text, graphics, and limited animation, as well as some degree of individualized interaction. The effect size of 0.39 standard deviations suggests, roughly, an improvement of 50th percentile students to the performance levels of 65th percentile students.

Interactive multimedia instruction involves more elaborate interactions adding more audio, more extensive animation, and (especially) video clips. The added cost of these capabilities may be compensated for by greater achievement—an average effect size of 0.50 standard deviations as compared with an effect size of 0.39 for garden-variety computer-based instruction. An effect size of 0.50 for interactive multimedia instruction suggests an improvement of 50th percentile students to the 69th percentile of performance.

Intelligent tutoring systems involve a capability that has been developing since the late 1960s (Carbonell, 1970), but has only recently been expand-

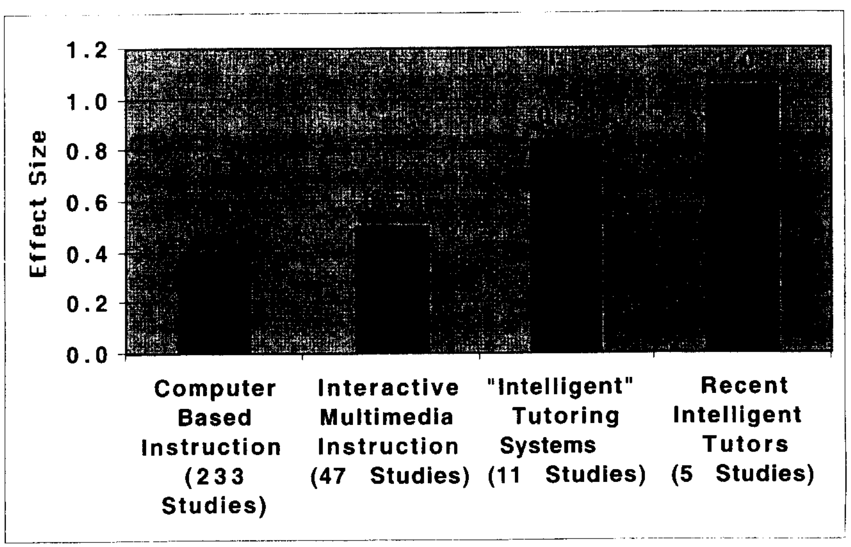

FIG. 4.1. Some effect sizes for technology-based instruction.

ing into general use. In this approach, an attempt is made to directly mimic the one-on-one dialogue that occurs in tutorial interactions. The key component is that computer presentations and responses are generated in real time, on demand, and as needed or requested by learners. Mixed initiative dialogue is supported in which either the computer or learner can ask or answer open-ended questions. Notably, these interactions are generated as required. Instructional designers do not need to anticipate and prestore them. This approach is computationally more sophisticated and it is more expensive to produce than standard computer-based instruction. However, its costs may be justified by the increase in average effect size to 0.84 standard deviations, which suggests, roughly, an improvement from 50th to 80th percentile performance.

Some later intelligent tutoring systems (Gott, Kane, & Lesgold, 1995) were considered just to see how far we are getting with this approach. The average effect size of 1.05 standard deviations for these recent applications is promising. It represents, roughly, an improvement of the performance of 50th percentile students to 85th percentile performance.

The more extensive tailoring of instruction to the needs of individual students that can be obtained through the use of generative, intelligent tutoring systems can only be expected to increase. They may raise the bar for the ultimate effectiveness of technology-based instruction. They may make available far greater efficiencies than we can now obtain from other approaches.

But further, these approaches may provide yet another example of what might be called the Columbus paradox. As readers will recall, Columbus sailed west to find India (and a lucrative spice route). Instead he opened the door to what became a new world for Europeans. Such a result typifies technological progress. Seeking one thing based on a metaphor with common practice, we almost inevitably end up with something at least unforeseen and often unexpected. Wireless telegraph produced something functionally quite different than the telegraph, namely, radio. Similarly, efforts to make a carriage run without a horse produced automobiles, to say nothing of gas stations, motels, and the Santa Monica Freeway. Seeking affordable one-on-one tutoring through automation, we may end up with something no one can envision at present. The existing metaphor is enough to get started, but the result may surprise us all.

Technology Can Be Used to Ensure That All Students Learn

One benefit of technology-based instruction appears to be that fewer students are left behind. Because sequence, pace, difficulty, content, and/or style of technology-based presentations can be tailored to each student's unique needs, some progress by each student can be ensured. Students can ported an overall effect size of 0.35, which is roughly equivalent to raising the achievement of 50th percentile students to that of 64th percentile students by exposing them to technology-based instruction.

Examination of results from meta-analyses concerning the use of technology-based instruction, tracking improvements in both technology and instructional approaches, produces the results shown in Fig. 4.1, which suggest progress toward Bloom's targeted effect size of two standard deviations.

In Fig. 4.1, computer-based instruction summarizes results from 233 studies that involved straightforward use of computer presentations using text, graphics, and limited animation, as well as some degree of individualized interaction. The effect size of 0.39 standard deviations suggests, roughly, an improvement of 50th percentile students to the performance levels of 65th percentile students.

Interactive multimedia instruction involves more elaborate interactions adding more audio, more extensive animation, and (especially) video clips. The added cost of these capabilities may be compensated for by greater achievement—an average effect size of 0.50 standard deviations as compared with an effect size of 0.39 for garden-variety computer-based instruction. An effect size of 0.50 for interactive multimedia instruction suggests an improvement of 50th percentile students to the 69th percentile of performance.

Intelligent tutoring systems involve a capability that has been developing since the late 1960s (Carbonell, 1970), but has only recently been expand-

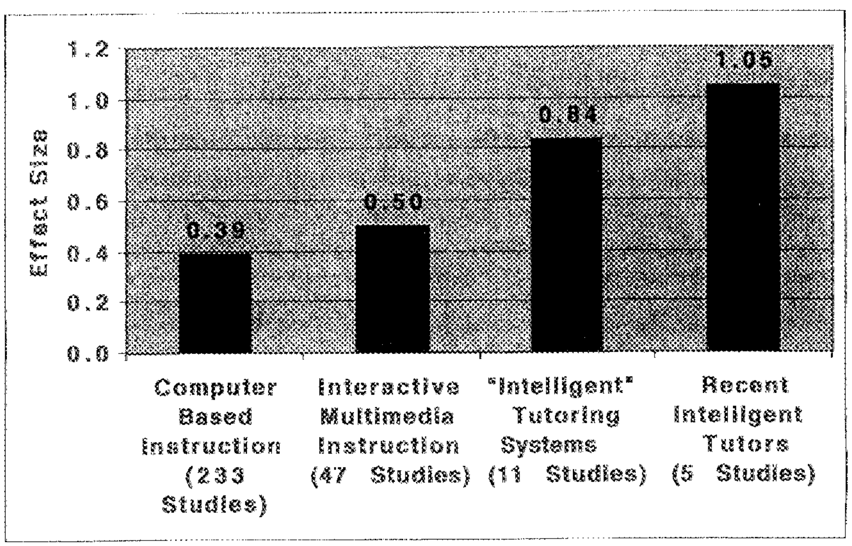

FIG. 4.1. Some effect sizes for technology-based instruction.

ing into general use. In this approach, an attempt is made to directly mimic the one-on-one dialogue that occurs in tutorial interactions. The key component is that computer presentations and responses are generated in real time, on demand, and as needed or requested by learners. Mixed initiative dialogue is supported in which either the computer or learner can ask or answer open-ended questions. Notably, these interactions are generated as required. Instructional designers do not need to anticipate and prestore them. This approach is computationally more sophisticated and it is more expensive to produce than standard computer-based instruction. However, its costs may be justified by the increase in average effect size to 0.84 standard deviations, which suggests, roughly, an improvement from 50th to 80th percentile performance.

Some later intelligent tutoring systems (Gott, Kane, & Lesgold, 1995) were considered just to see how far we are getting with this approach. The average effect size of 1.05 standard deviations for these recent applications is promising. It represents, roughly, an improvement of the performance of 50th percentile students to 85th percentile performance.

The more extensive tailoring of instruction to the needs of individual students that can be obtained through the use of generative, intelligent tutoring systems can only be expected to increase. They may raise the bar for the ultimate effectiveness of technology-based instruction. They may make available far greater efficiencies than we can now obtain from other approaches.

But further, these approaches may provide yet another example of what might be called the Columbus paradox. As readers will recall, Columbus sailed west to find India (and a lucrative spice route). Instead he opened the door to what became a new world for Europeans. Such a result typifies technological progress. Seeking one thing based on a metaphor with common practice, we almost inevitably end up with something at least unforeseen and often unexpected. Wireless telegraph produced something functionally quite different than the telegraph, namely, radio. Similarly, efforts to make a carriage run without a horse produced automobiles, to say nothing of gas stations, motels, and the Santa Monica Freeway. Seeking affordable one-on-one tutoring through automation, we may end up with something no one can envision at present. The existing metaphor is enough to get started, but the result may surprise us all.

Technology Can Be Used to Ensure That All Students Learn

One benefit of technology-based instruction appears to be that fewer students are left behind. Because sequence, pace, difficulty, content, and/or style of technology-based presentations can be tailored to each student's unique needs, some progress by each student can be ensured. Students can achieve the threshold of skill and knowledge required by the work for which they are preparing, thereby shortening the left tail of any distribution of achievement. This possibility is borne out by results. For example, Fletcher (1991) found in a review of 44 empirical comparisons, the spread (variance) of postinstruction measures, such as test scores or time to complete instruction, increased in every case more under conventional instruction than under interactive multimedia instruction. This trend was observed despite an overall increase in relative mean achievement scores of about 0.50 standard deviations under interactive multimedia instruction.

On the other hand, what happens if we open the gates and allow all students to progress as rapidly as they can? The differences noted in the pace with which students can proceed and the capacities of technology-assisted instruction for accommodating to these differences suggest that differences in achievement might well be expected to expand if instructional time is held constant.

Whatever the ultimate results, whether we ultimately decrease or increase the spread of achievement among age-based cohorts of students, instructional technology will help us deal with them. If students are allowed to stretch to their full learning potential, on one hand, and no student is lost in the rush to learn, on the other, we will increase fairness and opportunity equity in instruction at both ends of the spectrum.

Technology Can Be Used to Reduce Time Needed to Reach Instructional Objectives

If instructional time is not spent re-presenting material the student already knows and is concentrated on material the student has yet to learn, learning should occur more quickly. As suggested by Table 4.1, research suggests that it does.

This finding arises repeatedly in reviews of instructional technology. Orlansky and String (1977) found that reductions in time to reach instructional objectives averaged about 54% in their review of CBI used in military training. Fletcher (1991) found an average time reduction of 31% in 6 studies of interactive videodisc instruction applied in higher education. Kulik

TABLE 4.1
Percent Time Savings for Technology-Based Instruction

Study (Reference)	*Number of Findings*	*Average Time Saved (%)*
Orlansky and String (1977)	13	54
Fletcher (1991)	6	31
Kulik (1994) (Higher Education)	17	34
Kulik (1994) (Adult Education)	15	24

and his colleagues found time reductions of 34% in 17 studies of CBI used in higher education, and 24% in 15 studies of adult education (Kulik, 1994). These reviews are effectively independent in that they reviewed different sets of evaluation studies. On this basis, reductions of about 30% in the time it takes students to reach a variety of given instructional objectives seem to be a reasonable expectation.

It turns out that 30% is a fairly conservative target. Commercial enterprises that develop technology-based instruction for the Department of Defense (DoD) regularly base their bids on the expectation that they can reduce instructional time by 50%. Noja (1991) reported time savings through the use of technology-based instruction as high as 80% in training operators and maintenance technicians for the Italian Air Force.

Time savings of 30% are not an inconsiderable matter for the Department of Defense. The DoD spends about $4 billion a year on specialized skill training, which is the postbasic training needed to qualify people for the many technical jobs (e.g., wheeled vehicle mechanics, radar operators and technicians, medical technicians) needed to perform military operations. If the DoD were to reduce time to train 20% of the people undergoing specialized skill training by 30%, it would save over $250 million per year. If it were to do so for 60% of the people undergoing specialized skill training, it would save over $700 million per year.

It is harder to assign dollar values to the time that students spend in educational settings, especially K–12 classrooms, but time so spent is not without cost and value. Aside from the obvious motivational issues of keeping students interested and involved in educational material, using their time well will profit both them and any society that depends on their eventual competency and achievement. The time savings offered by technology-based instruction in K–12 education may be more significant and of greater value than those obtained in posteducation training.

Students Prefer Technology-Based Instruction

Many evaluations of technology-based instruction simply ask students if they prefer it to more conventional classroom approaches. Greiner (1991) reviewed these evaluations and found that typically from 70% to 80% of students polled preferred technology-based approaches over those that were not technology based. When students reported that they did not prefer such approaches, the reasons were usually traced to implementation or technical problems with the technology, not the instructional approach itself.

McKinnon, Nolan, and Sinclair (2000) completed a thorough 3-year study of student attitudes toward the use of computers as learning and productivity tools for such applications as spreadsheets, databases, graphics, desktop publishing, and statistical processing in their junior high school English, mathematics, science, and social science courses. The academic performance of the computer-using students remained steadily and significantly superior to that of their noncomputer-using peers throughout the 3 years of the study.

As might be expected, the attitudes of the computer-using students in this study toward computer use slackened and became less positive as the novelty of using the technology wore off. However the attitudes of the computer-using students remained positive and significantly more positive than those of the noncomputer using students throughout all 3 years of the study. After examining their data, McKinnon et al. attributed the decrease in positive attitudes among the computer-using students to "habituation" in using computers rather than any disenchantment with their utility. Computers simply became tools for classroom learning that the students took for granted.

Technology-Based Instruction Can Be Cost-Effective

The central question for researchers often concerns whether or not a new approach is an improvement over current practice. The central question for educational decision-makers, however, must go a step farther. Decision-makers must, of course, be concerned with improving current educational practice, but they must also consider what must be given up to do so. In the practical world, this usually leads to consideration of costs as well as effectiveness. If researchers wish to make a difference in educational practice, they must address decisions of cost-effectiveness.

Cost-effectiveness may be assessed in either of two ways. We may either hold costs constant and assess alternatives for maximizing effectiveness, or we may hold effectiveness constant and assess alternatives for minimizing costs. Adequate models of both costs and effectiveness must be employed in studies using either of these approaches. Few studies using either approach have been performed in education, and more are needed.

One comparison holding effectiveness constant while seeking minimized costs is shown in Table 4.2. The table uses empirical data reported by Jamison, Fletcher, Suppes, and Atkinson (1976), Levin, Glass, and Meister (1987), and Fletcher, Hawley, and Piele (1990). It reports the costs (in constant 2001 dollars) to raise comprehensive mathematics scores on a standardized test one standard deviation using several different instructional approaches: tutors, reduced class size, increased instructional time, and computer-based instruction.

The table suggests that the most cost-effective approaches among all these alternatives are computer-based instruction and peer tutoring. It also suggests that, of the two, computer-based instruction is more cost-effective.

TABLE 4.2
Costs (Constant $2001) to Raise Mathematics Scores by One Standard Deviation

Alternative	*Costs*
Tutoring (20 min/day):	
Peer tutors	472
Adult tutors	2,654
Reduce class size from:	
35 to 30	1,619
35 to 20	2,251
Increase instruction time 30 min/day	4,391
Computer-based instruction for 10 min/day:	
- Mini-computer in laboratory (1976)	
Grade 3	382
Grade 5	785
- Microcomputers in classrooms (1990)	
Grade 3	316
Grade 5	339

This result echoes the findings of Niemiec, Sikorski, and Walberg (1989), who compared studies of the costs and effectiveness of peer tutoring with studies of computer-based instruction. They found the two approaches to be equally effective and both to be more effective by about 0.4 standard deviations than conventional classroom instruction. They also found a clear cost-effectiveness superiority (by a factor of about three) for computer-based instruction over peer tutoring.

It is notable that these two approaches are not incompatible and that a strong cost-effectiveness argument can be made for combining peer tutoring with computer-based instruction. Such a combination may be accomplished by presenting instruction to more than one student at a time on a single computer station. This sort of approach has been shown early on by Grubb (1964), and more recently by Shlecter (1990), to be effective.

CAVEATS

A summary of research results must necessarily slide over many issues of intent, design, implementation, and evaluation. These issues may also be summarized.

Assessment of Innovation

One difficulty for any evaluation of an innovative technology is that there is nothing else like it. Each educational approach has its own strengths and limitations. If an evaluation is held to strict instructional and experimental controls based on constraints imposed by one approach, then the other approach will be at a disadvantage. More specifically, in comparisons of older with newer approaches, the older approach may limit application, and assessment, of the newer approach.

New approaches are often not used to best advantage because not enough is understood about how best to use them, they have not matured sufficiently for fair comparison with what's currently available, and/or sufficient infrastructure does not exist to support them adequately. Early horseless carriages were certainly inferior in both costs and effectiveness to horse-drawn carriages if they were viewed strictly as a means for getting from one place to another. They were misused by their drivers, their design and production had not matured sufficiently to make them reliable, and the necessary infrastructure of roads, gas stations, competent mechanics, and so on did not exist. The promise of many instructional technologies and their applications may similarly be masked by our unfamiliarity with them. Despite all the evaluations discussed, instructional approaches that use these technologies to best advantage may not yet exist.

Single Factor Assessment

Empirical evaluations are always subject to one experimental contamination or another. It is rare (i.e., never) that we find comparisons in which the single different factor is the presence or absence of instructional technology. For instance, the instructional content and objectives of the original approach may be revised and incorporated in the new, but not the original, approach. The revised body of materials may then be compared with the original instruction and owe more of its observed success to the revision than to the functionalities that are the object of the evaluation. Problems such as this seem unavoidable. The "trajectory theory" approaches of Suppes et al. (1975, 1976) may provide a way around this problem. Using only the amount of time a new approach is in active use to predict achievement on external measures, such as standard test scores, we may get as close to single factor assessment as we can in field research and assessments.

On the other hand, if we can not be perfect, at least we can be explicit. Careful documentation will ensure that everyone knows what has been done and can weigh conclusions for themselves. Seidel and Perez (1994), many of Reigeluth's (1999) contributors, and others supported this position by emphasizing that candidate innovations for instructional practice must consider the full environment into which an innovation is placed. More attention to full system environments may be needed in evaluations of technology-based instruction. Even when these environments are carefully considered, comprehensive discussions of them are rarely presented in evaluation research literature, which is often constrained by such mundane considerations as publication space. Meta-analytic reviews like those discussed above may provide a partial solution to this problem. If it is impossible to control for all factors, perhaps we can perform and then review enough assessments to ensure that the signal can be separated from the noise of extraneous factors.

Production Quality

Aspects such as the quality of graphics, clarity of instructional text, verisimilitude of simulations, and relevance of tutorial advice may have a substantial impact on the effectiveness of many instructional technology applications, but these issues are rarely addressed. The impact of production quality, especially the impact of its costs on resulting instructional effectiveness, needs to be better understood. It may explain an appreciable portion of the difference in otherwise similar approaches.

Media-Based Assessment

No discussion of technology applications in instruction would be complete without mentioning Clark's (1983) argument that "the best current evidence is that media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition" (p. 445). These concerns may be summed up by the notion that technology alone does not define an instructional approach: What is done with the technology is what counts. This point of view seems unequivocal. The presence of technology is no guarantee that effective instructional content, effective ways to present it, or even that the unique strengths of the technology will be used.

On the other hand, the absence of technology is a reasonable guarantee that its unique capabilities will be missing. It is difficult to imagine approaches that make the intense individualization discussed above affordable or even possible without technology. To take Clark's metaphor directly, improvements in the technology of delivering food from centers of production to markets have had a tremendous impact on the nutrition of nations. The technologies by themselves do not guarantee any impact, but the functionalities they make possible put serious improvements within reach, if not our grasp.

Moreover, arguments such as Clark's are often advanced as blanket recommendations of what should or should not be considered in evaluation research. However, evaluation is performed either explicitly or implicitly to inform a decision. Should we do A or B? How might we improve C? What should our policy be toward D? Should we invest our scarce resources in E? Should we institutionalize the use of F? Aside from technical and methodological quality, blanket prescriptions for what ends an evaluation should or should not serve seem misguided.

But what then should we make of the above meta-analyses of technology-based instruction? What decisions might they inform? These analyses review what has been observed to happen when technology is applied in instruction. They do not address cause and effect, nor do they address how either the technology or the full learning environment should be designed to promote learning or achievement of specific learning outcomes. The state of the art does not seem ready to address these issues, although many individual studies hint at significant possibilities.

These analyses do report data now available from systematic, reasonably well-designed and implemented evaluation studies, and they suggest reasons why we might well expect these data to be favorable. Technology can take on the overwhelming human workload that classroom teachers would need to assume in order to achieve significant new levels of effectiveness in education and to prepare students for the massive, global infusion of technology in the adult working world. If these are our goals, then available data suggest that technology can make their accomplishment both affordable and accessible.

If a decision must be made concerning future investment in these technologies, then the task of the evaluator is to inform the decision as well as possible. Based on the data reviewed above, it seems reasonable, if not imperative, to proceed. Despite the great amount of work that must be done to learn how best to use technology in education, "paralysis by analysis" does not seem to be an acceptable option.

Third-Party Evaluation

Many evaluations of instructional technology applications are performed by their developers or others who have a stake in their success. There are both strengths and weaknesses to such evaluations. Developers are rarely indifferent to the success of their products and may, intentionally or not, bias the results of their evaluation. On the other hand, they also have a stake in honest assessment, and they may understand better than anyone the strengths and limitations of what they have produced. Despite current emphases on the virtues of third-party evaluations, such evaluations should not be sought without question, and evaluations performed by developers and other stakeholders should not be discarded as irrelevant. We should seek to understand better the implications of both sorts of evaluation and the data, findings, and information they provide. Many of the evaluation studies reviewed in the above meta-analyses were performed by developers rather than third parties. Based on the methodological filters applied in selecting

the studies and discussions of their results, there seem to be many good reasons to judge their findings as valid.

Antiquity of Applications

By the time an evaluation study is performed, documented, and reported in a form accessible to developers and potential users, the application originally under consideration is likely to be 5 or more years old. By the time it is included in a summary such as this one, the application may be well over 10 years old (or more) and superseded by curriculum requirements, job design, or many other factors that change over time. It is unlikely that the application itself will be of more than historical interest to decision-makers.

However, the principles underlying the design of the applications and their success may well be of continuing interest to designers and potential users. The technologies and media employed in the application are likely to be notable as long as they receive investment. As usual, the value and relevance of the evaluation do not depend as much on the antiquity of the application being considered as on how well it informs the decisions that will be made based on its findings. In a world of rapidly emerging technology and applications, useful information on a specific application is frequently unavailable until the state of the art passes it by. However, the class of applications to which it belongs may persist long enough to be usefully informed by assessments such as those discussed above.

Relation of Instructional Design to Outcomes

Different outcomes, or instructional objectives, must compete for scarce instructional resources. Decisions made in the design of instructional technology programs affect both their costs and their achievement of specific instructional objectives. These relationships should be better understood. How, for instance, should a program be designed to maximize transfer ability, retention, speed of response, accuracy of response, or motivation to continue study? What do the design alternatives cost? To what quantitative degree do they contribute to instructional effectiveness? How should we trade them off against one another, as we invariably must in the practical world of training design?

The individualization of control and student progress that can be exercised in instructional technology applications, raises these issues to a level of both significance and practical payoff that they do not reach elsewhere. Technology offers a degree of control that makes serious engineering of instruction practicable. Raymond Fox (1994) stated that "one of the more difficult problems in dealing with improvement in public education is to replace the notion of teaching as an art form with that of instruction delivery as a systems science" (p. 2). If we are to achieve an engineering of instruction so that predictable results can be reliably obtained by many hands, we must seek an engineering of instructional design that provides predictable outcomes from specific choices among design alternatives. We must seek to learn not only if instructional technology works, but also how it works.

CONCLUSIONS

The results discussed above suggest that applications of technology in education and training are more effective than our current practice. To conclude that they are not would require a substantial flood of findings to the contrary. On the other hand, there remain significant questions to be answered.

How can new technologies best be integrated with existing institutions of instruction? Most change is evolutionary rather than revolutionary, and technology-based approaches to instruction seem no different, despite the immense changes they may ultimately engender. In a constructivist world, we cannot make learning occur, we can only create environments that promote and encourage it (e.g., Mayer, 1999). With or without constructivism, Seidel and Perez (1994) were right. More needs to be done to understand how best to design and implement these environments. The revolution wrought by the horseless carriage was not complete until the infrastructure it required was put into place. The same will be true of technology-based instruction.

Costs must be considered as well as effectiveness. Educational researchers who dedicate their professional lives to improving education by developing new and innovative approaches are well advised to examine effectiveness. Most of them rise responsibly and well to the challenge. For educational decision-makers, however, effectiveness may be only half of the question. Somewhere costs must be considered by someone. Proper consideration of costs depends on the development of adequate models of costs and detailed understanding of the cost consequences arising from any instructional innovation must be developed, especially an innovation involving such an alien approach as technology-based instruction. Few educational decision-makers have time or resources to attend to these matters, and they, like most of us, have a vested interest in maintaining the status quo. That leaves educational researchers to fill the gap. Just as they rise to the challenge in evaluating new approaches to instruction, so they must also begin to address seriously the cost consequences of the new approaches they choose to champion.

How should we design technology-based instruction? The state of the art and practice in technology-based instruction has passed beyond questions

of *if* applications using it work to *how* they work. How might instruction and instructional enviroments be designed to bring about specific instructional outcomes? Excellent reviews by Krendl and Lieberman (1988) and Schater and Fagnano (1999) echo earlier recommendations by Suppes (1964) and others to apply advances in cognitive and learning theory to the development of technology-based instruction. Such efforts will both improve the quality of instruction delivered and, probably more importantly, provide feedback to theories of cognition and learning about where they are right, where they might use some improvement, and where they have left gaps that badly need to be filled. This is the traditional interplay of theory and empirical research that has served other areas of systematic investigation so well. Technology-based instruction, with its precise control over inputs and equally precise measurement of outputs, appears to have a unique role to serve in completing the feedback loop between instructional theory and instructional research.

How might we best individualize instruction? If individualization is the key enabling capability of technology-based instruction, then more should be done to determine how best to accomplish it. What cues should be gathered from the interactions of individuals with technology to best tailor sequence, style, content, and pace of the instructional presentations for them? There are bright spots in the research literature. More could and should be done to identify, learn from, and develop them.

This review of technology-based instruction suggests that it will most probably lower costs and increase effectiveness for many applications. It is likely to emerge as the most cost-effective alternative in many settings and applications when considered among all other possibilities in a full systems context. Overall, it is likely to improve both our practice of instruction and theories of learning and cognition. It does not seem unreasonable, then, to argue that the resources needed to realize its potential are well spent. These resources include funding, time, and the effort to effect significant changes in professional practice and in our instructional institutions. The point of the above review has been to suggest that the return on this investment will be sizable and worthwhile.

REFERENCES

Allan, D. M. E. (1989). *The effectiveness of interactive videodisc in clinical medical education.* San Diego, CA: Intelligent Images, DAROX Corporation.

Atkinson, R. C., & Wilson, H. A. (Eds.). (1969). *Computer-assisted instruction.* New York: Academic Press.

Bitzer, D. L., Braunfeld, P. G., & Lichtenberger, W. W. (1962). Plato II: A multiple-student, computer-controlled, automatic teaching device. In J. E. Coulson (Ed.), *Programmed learning and computer-based instruction* (pp. 205–216). New York: Wiley.

Bloom, B. S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Carbonell, J. R. (1970). AI in CAI: An artificial intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems, 11*, 190–202.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*, 445–459.

Crotty, J. M. (1984). *Instruction via an intelligent videodisc system versus classroom instruction for beginning college French students: A comparative experiment.* (AFIT/CI/NR 84-51D). Wright-Patterson AFB, OH: Air Force Institute of Technology/NR. (DTIC No. ADA 145 599). (University Microfilms No. 8513797).

Fletcher, J. D. (1979). The design of computer-assisted instruction in beginning reading: The Stanford projects. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading instruction* (pp. 243–267). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fletcher, J. D. (1991). Effectiveness and cost of interactive videodisc instruction. *Machine Mediated Learning, 3*, 361–385.

Fletcher, J. D. (1997). What have we learned about computer based instruction in military training? In R. J. Seidel & P. R. Chatelier (Eds.), *Virtual reality, training's future?* (pp. 169–177). New York: Plenum.

Fletcher, J. D., Hawley, D. E., & Piele, P. K. (1990). Costs, effects, and utility of microcomputer assisted instruction in the classroom. *American Educational Research Journal, 27*, 783–806.

Fox, R. G. (1994). *A systems approach to education.* Warrenton, VA: Learning Technology Institute.

Gettinger, M. (1984). Individual differences in time needed for learning: A review of the literature. *Educational Psychologist, 19*, 15–29.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3–8.

Gott, S. P., Kane, R. S., & Lesgold, A. (1995). *Tutoring for Transfer of Technical Competence* (AL/HR-TP-1995-0002). Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104–137.

Greiner, J. M. (1991). Interactive multimedia instruction: What do the numbers show? In *Proceedings of the Ninth Annual Conference on Interactive Instruction Delivery* (pp. 100–104). Warrenton, VA: Society for Applied Learning Technology.

Grubb, R. E. (1964). *The effects of paired student interaction in the computer tutoring of statistics.* Yorktown Heights, NY: Thomas J. Watson Research Center.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Heuston, D. H. (1997). *School improvement models: The manual model and the speed of light.* Sandy, UT: Waterford Institute.

Holland, J. G. (1959). A teaching machine program in psychology. In E. Galanter (Ed.), *Automatic teaching: The state of the art* (pp. 69–82). New York: Wiley.

Jamison, D. T., Fletcher, J. D., Suppes, P., & Atkinson, R. C. (1976). Cost and performance of computer-assisted instruction for education of disadvantaged children. In J. T. Froomkin, D. T. Jamison, & R. Radner (Eds.), *Education as an industry* (pp. 201–240). Cambridge, MA: Ballinger.

Krendl, K. A., & Lieberman, D. A. (1988). Computers and learning: A review of recent research. *Journal of Educational Computing Research, 4*, 367–389.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in education and training* (pp. 9–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Levin, H. M., Glass, G. V., & Meister, G. R. (1987). Cost-effectiveness of computer-assisted instruction. *Evaluation Review, 11*, 50–71.

Mayer, R. E. (1999). Designing instruction for constructivist learning. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: Vol. II. A new paradigm of instructional theory* (pp. 141–159). Hillsdale, NJ: Lawrence Erlbaum Associates.

McKinnon, D. H., Nolan, C. J. P., & Sinclair, K. E. (2000). A longitudinal study of student attitudes toward computers: Resolving an attitude decay paradox. *Journal of Research on Computing in Education, 32*, 325–335.

Niemiec, R. P., Sikorski, M., & Walberg, H. J. (1989). Comparing the cost-effectiveness of tutoring and computer-based instruction. *Journal of Educational Computing Research, 5*, 395–407.

Noja, G. P. (1991). DVI and system integration: A further step in ICAI/IMS technology. In R. J. Seidel & P. R. Chatelier (Eds.), *Advanced technologies applied to training design* (pp. 161–189). New York: Plenum.

Orlansky, J., & String, J. (1977). *Cost effectiveness of computer-based instruction in military training* (IDA Paper P-1375). Arlington, VA: Institute for Defense Analyses.

Porter, D. (1959). Some effects of year long teaching machine instruction. In E. Galanter (Ed.), *Automatic teaching: The state of the art* (pp. 85–90). New York: Wiley.

Reigeluth, C. (1999). *Instructional design, theories, and models: A new paradigm of instructional theory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schacter, J., & Fagnano, C. (1999). Does computer technology improve student learning and achievement? How, when, and under what conditions. *Journal of Educational Computing Research, 20*, 329–343.

Scriven, M. (1975). Problems and prospects for individualization. In H. Talmage (Ed.), *Systems of individualized education* (pp. 199–210). Berkeley, CA: McCutchan.

Seidel, R. J., & Perez, R. S. (1994). An evaluation model for investigating the impact of innovative educational technology. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in software applications* (pp. 177–212). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shlechter, T. M. (1990). The relative instructional efficiency of small group computer-based training. *Journal of Educational Computing Research, 6*, 329–341.

Suppes, P. (1964). Modern learning theory and the elementary-school curriculum. *American Educational Research Journal, 1*, 79–93.

Suppes, P., Fletcher, J. D., & Zanotti, M. (1975). Performance models of American Indian students on computer-assisted instruction in elementary mathematics. *Instructional Science, 4*, 303–313.

Suppes, P., Fletcher, J. D., & Zanotti, M. (1976). Models of individual trajectories in computer-assisted instruction for deaf students. *Journal of Educational Psychology, 68*, 117–127.

Tobias, S. (1982). When do instructional methods make a difference? *Educational Researcher, 11*(4), 4–9.

Uttal, W. R. (1962). On conversational interaction. In J. E. Coulson (Ed.), *Programmed learning and computer-based instruction* (pp. 171–190). New York: Wiley.

Verano, M. (1987). *Achievement and retention of Spanish presented via videodisc in linear, segmented, and interactive modes.* (Doctoral dissertation, University of Texas, Austin, TX) (DTIC No. ADA 185 893).

Vinsonhaler, J. F., & Bass, R. K. (1972). A summary of ten major studies on CAI drill and practice. *Educational Technology, 12*, 29–32.